SPATIAL SCALABILITY

FOR FINE GRANULAR VIDEO ENCODING

BENEFIT OF EARILIER FILING DATE

[0001] This application claims the benefit of provisional patent application serial no. 60/239,347, entitled "SPATIAL SCALABILITY WITH FGS, filed on October 11, 2000, the entirety of which is hereby incorporated by reference.

RELATED APPLICATION

[0002] This application is related to commonly assigned: U.S. Patent Application Serial Number _____, entitled " Totally Embedded FGS Video Coding with Motion Compensation", filed _____; U.S. Patent Application Serial Number _____, entitled "Double Loop Fine Granular Scalability", filed on _____; U.S. Patent Application Serial Number _____, entitled "Single Loop Fine Granular Scalability", filed on _____.

FIELD OF THE INVENTION

[0003] This invention relates generally to video encoding and more specifically to spatial, or resolution, scalability for fine granular scalability encoded video signals.

BACKGROUND OF THE INVENTION

[0004] The flexibility of a Fine-Granular Scalability (FGS) video encoding enables it to support a wide range of transmission bandwidths as is described in U.S. Patent Application Serial Number _____, entitled "System and Method for Improved Fine Granular Scalable Video Using Base Layer Coding Information" filed on _____, and assigned to the assignee herein. An improved FGS video encoding method is disclosed in U.S. Patent Application Serial Number _____, entitled "Hybrid Temporal-SNR Fine Granular Scalability Video Coding," filed on _____ and assigned to the assignee herein. In the hybrid temporal-SNR FGS video encoding method disclosed, an encoder is operable to encode and transmit video images with improved quality, referred to herein as SNR, FGS encoded form, or with improved temporal form or in a combined hybrid temporal-SNR FGS form.

[0005] However, in the current FGS framework there is no ability to alter the image resolution (i.e., spatial scalability) to conform to different resolutions, such as QCIF or CIF. The addition of spatial scalability to FGS encoding would be advantageous as the current FGS encoding system limits the transmission bandwidth range due to the predefined spatial resolution at encoding time. Since variable bandwidth networks, such as the Internet, have bandwidths, or bit rates, that vary widely and are generally unknown at encoding time, the current system does not adapt to the varying bandwidth and, hence, is not adequate for variable bandwidth network video transmission.

[0006] Hence, there is a need for a spatial scalability structure with the current FGS encoding system that allows for a wider range of transmission bit rates to be satisfied.

SUMMARY OF THE INVENTION

[0007] A method for coding video data to allow for scalability of image resolution comprising downscaling the video data image frames, encoding the downscaled video data to produce base layer frames, generating quality enhanced residual images from the downscaled video data and said encoded data in said base layer frame, encoding said quality enhanced residual images using a fine granular coding technique to produce quality enhancement frames, upscaling the encoded data contained in the base layer frames and corresponding quality enhanced residual frames, generating a first set of residual images from the upscaled base layer, and corresponding quality enhancement layer frame data and the video data, encoding this first set of residual images using a fine granular coding technique to produce spatial enhancement frames. In another aspect of the invention, a second temporal enhancement frame information is produced from among the information contained in the encoded spatial enhancement layer frames.

BRIEF DESCRIPTON OF THE FIGURES

[0008] Figure 1 depicts a conventional FGS encoding system;

[0009] Figure 2 illustrates hybrid temporal-SNR FGS video encoded structures;

[0010] Figure 3 illustrates an exemplary video image encoding process in accordance with the principles of the invention;

[0011] Figure 4a illustrates an exemplary spatial encoder in accordance with the principles of the invention;

[0012] Figure 4b illustrates a second exemplary spatial encoder in accordance with the principles of the invention;

[0013] Figure 5 illustrates a first exemplary spatial-SNR FGS video encoded structure developed using the encoding system illustrated in Figure 4b;

[0014] Figure 6 illustrates a second exemplary spatial-SNR FGS video encoded structure developed using the encoding system illustrated in Figure 4b;

[0015] Figure 7 illustrates an exemplary hybrid spatial-temporal-SNR FGS video encode structure developed using the encoding system illustrated in Figure 4b;

[0016] Figure 8 illustrates a second exemplary hybrid spatial-temporal SNR FGS video encoded structure;

[0017] Figures 9a illustrates an exemplary system configuration that utilizes the principles of the present invention;

[0018] Figures 9b illustrates a second exemplary system configuration that utilizes the principles of the present invention; and

[0019] Figure 10 illustrates an exemplary block diagram of a transmission system incorporating an encoding system in accordance with the principles of the present invention.

[0020] It is to be understood that these drawings are solely for purposes of illustrating the concepts of the invention and are not intended as a level of the limits of the invention. It will be appreciated that the same reference numerals, possibly supplemented with reference characters where appropriate, have been used throughout to identify corresponding parts.

DETAILED DESCRIPTION OF THE INVENTION

[0021] Figure 1 illustrates system 100 for encoding video images in a hybrid temporal-SNR FGS encoding structure. System 100 receives video images from video source 2 and transmits encoded video images across variable bandwidth network 6. As will be appreciated, video source 2 can be embodied by any type of video capturing device, such as television camera, video recorder/playback, analog or digital, etc., and the variable bandwidth network, may be a landline network, such as the Internet, a point-to-point network, such as the telephone network, or a wireless network, such as a satellite channel or a mobile receiving device, such as a cellular phone or computer.

[0022] Encoder 110 is composed principally of a base layer (BL) encoder 8, a hybrid temporal-SNR FGS video encoder 20 and video rate controller 18. Base layer encoder 8, which is described in the earlier referenced application, Serial Number _____, encodes received video images into a base layer data stream. The encoded base layer represents a level of encoding that is representative of a minimally acceptable video image and is guaranteed to be transmitted over network 6. FGS layer encoder 20, which is described in the earlier referenced application, Serial Number _____, encodes residual images generated between the input video images and base layer encoded images of the input video images into a video enhancement layer. The video enhancement layer is used to improve the quality of an image produced by the encoded base layer. Rate controller 18 determines the rate of transmission of the base layer and enhancement layer, and consequently the number of bits that can be transmitted, depending upon, for example,

available bandwidth and user preference. User preference can be input to controller 18 by user input 3.

[0023] As illustrated, video data from video source 2 is input to both BL encoder 8 and hybrid temporal-SNR FGS video encoder 20. BL encoder 8 encodes an original video image using a conventional frame-prediction coding technique and compresses the video data at a predetermined bit-rate, represented as R_{BL}. Calculation block 4 sets R_{BL} to a value between a minimum bit rate (R_{min}) and a maximum bit rate (R). In most cases R_{BL} is set to Rmin to ensure even at lowest bandwidths, network 6 will be able to accommodate the video data coded by base layer encoder 8.

[0024] The original video data from source 2 and the coded video data (i.e., base layer encoded image) provided by BL encoder 8 are further provided to both a residual image (RI) computation block 10 and motion compensated residual image (MCRI) computation block 24 in hybrid encoder 20. RI computation block 10 and MCRI computation block 24 process the original video data and the coded video data to generate residual images 12 and motion compensated (MC) residual images 22, respectively. Residual images 12 are generated based on a difference between the pixels in this decoded video data and the corresponding pixels in the original video data. The MCRI computation block 24 receives coded video data from BL encoder 8 and also decodes this encoded video data. The MC residual images 22 are generated based on a motion-compensation approach from the decoded video data.

[0025] As a result of the above hybrid coding, two streams of enhancement layer frames are produced; a temporal enhancement stream 32, referred to herein as FGST encoded, and an enhancement stream 31, referred to herein as FGS encoded. The FGST

encoded enhancement steam 32 includes the compressed FGS temporal frames from the MCRI EL encoder 26 while the FGS encoded enhancement stream 31 includes the SNR, i.e., standard FGS residual, frames from residual image encoder 14. Video encoded streams 31, 32 can be transmitted independently or combined to produce a single enhancement layer stream.

[0026] Figure 2 illustrates one exemplary example of a hybrid temporal-SNR FGS scalability structure achievable with encoder 110 illustrated in Figure 1. A base layer 210 includes, as "I" frames and "P" frames, which are represented in this illustrative example, as "I" frame 212, and "P" frames, 214, 216, 218, etc. "I" frame 212 is representative of encoded video image data and "P" frames 214, 216, 218, etc., are representative of predicated data frames. In this illustrated example, there is a single "I" frame for a plurality of "P" frames. However, it will be appreciated at that the number "I" frames and "P" frames may vary as the content of the encoded images varies.

[0027] Also illustrated are quality (FGS) enhancement layer 240 and temporal (FGST) enhancement layer 230, which, as previously described, are used to achieve quality and temporal scalability, respectively, of an original video image. In this case, temporal FGST layer 230 is used to add temporal enhancement to the encoded information in "I" frame 212 and "P" frames 214, 216, 218, etc., contained in base-layer 210 and FGS layer 240 is used to add quality improvement to the encoded video images in base layer 210 frames and temporal enhancement layer 240 frames.

[0028] In this illustrative example, the encoded "I" frame 212 and "P" frames 214, 216, 218, etc., contained in base layer 210. “P” frames 214, 216, 218, etc. contain residual information between the original video image and “I” frame 212. FGS enhancement layer

240 and FGST enhancement layer 230 frames are further shown in an alternating sequence. In this case, the frame data in FGST layer block 232, for example, are predicted from encoded data in base layer frames 212, and 214, as is well known in the art. In this case, frames within base layer 210 and FGST layer 230 are transmitted in alternating manner.

[0029] Figure 3 illustrates an exemplary processing flow 300 in accordance with the principles of the invention. In this exemplary flow, video images 2' are representative of high-resolution images. Video images 2' are first downscaled using well known a downscaling process 310, to a much lower resolution. The downscaled video images 315 are then encoded into a base layer 320 and an enhancement layer 330 using FGS encoding. More specifically, downscaled video image 2' is encoded into base layer 320, which is representative of a minimally acceptable video image. Residual layer 330 is then created from the original downscaled image 315 and corresponding image base layer 320.

[0030] The original image is next reconstructed by first upscaling the encoded base layer and enhancement layer using a well known upscale process 340. A residual enhancement layer is then created using the reconstructed-upscaled video image and the original video image 2'. The residual is FGS encoded and is used to adapt the image transmission to satisfy varying network bandwidth and resolution constraints.

[0031] Figure 4a illustrates a block diagram of a spatial FGS encoding system in accordance with the principles of the invention. In this encoding system high-resolution video images 2' from video source 2 are provided to resolution downscaler 410, and spatial encoder 430. Method used in downscaler 410 for downscaling high-resolution image are well known in the art. For example, high-resolution images of 1024x768 pixels may be downscaled by 352x488 pixels, i.e., CIF format, by selecting every third pixel row and

column. Another technique may average the pixels within a square matrix block of a known number of pixels, e.g., 3x3 matrix.

[0032] The downscaled image is next provided to a base layer encoder 8 to generate a downscaled base layer. The downscaled base layer and a reconstructed downscaled image are combined using summer 415 and provided to enhancement encoder 420. Enhancement layer encoder 420 determines a quality, a temporal, a spatial or a combined quality/temporal/spatial enhancement layer. The output of encoder 420 is represented as the SNR stream 30 and spatial stream 440.

[0033] A reconstructed image based on the base layer/quality enhancement layer image is produced by summer 425 and upscaler 430. Subtractor 435 then determines the difference, i.e., residual, between original image 2' and the reconstructed image based on the upscaled base/quality enhancement layer image. The residual of the upscaled image base/quality layer and the output of summer 415 are provided to formatter 437 for formatting into bitplanes. The images, formatted into bit planes, are input to FGS encoder 439 to transmission as SNR enhancement layer 30 or spatial enhancement layer 440 over network 6.

[0034] Figure 4b illustrates a second exemplary encoder in accordance with the principles of the present invention. In this embodiment a second subtractor 460 is incorporated into temporal/spatial encoder 420 to determine another residual layer that is representative of a temporal layer of a spatial layer. It would be understood by those skilled in the art that the operation of the encoder illustrated in Figure 4b is similar to that illustrated in Figure 4a.

[0035] Figure 5 illustrates an exemplary embodiment of spatial scalability structural in accordance with the principles of the encoder illustrated in Figure 4a. In this embodiment,

original video signal 2' is first downscaled to a lower resolution, or image size, and encoded in a base layer 210' using known encoding methods, preferably FGS encoding. In this case, base layer frames 212', 214', 216', etc., are coded up to a bit rate (or bandwidth) R_{Bl}, which is less than the minimum bit rate, R_{min}. SNR enhancement layer 240' frames 242', 244', 246', etc. are generated from the residual signal of the downscaled video image and base layer frames. SNR enhancement frames are coded until a known available bandwidth, R'.

[0036] A high resolution enhancement layer 510, i.e., spatial layer, is then created from the upscaled low resolution base layer 210' and a portion of the enhancement layer 240' and original high resolution image 2'. Spatial layer frames 512, 514, 516, etc. are encoded for the remainder of the maximum available bandwidth, R_{max}. In an alternate embodiment layer 510 could directly predicted from base layer 210'.

[0037] The spatial scalability illustrated in Figure 5 is advantageous as it can extend the range of transmission bit rates since various resolutions can be transmitted. Thus, standard FGS encoding can encode video images with a bit rate in the interval R_{min} to R', as previously discussed, and spatial scalability can extend the bit rate to R_{max} from R_{min0}. Spatial scalability can reduce the minimum bandwidth as the downscaled base layer has a reduced number of pixels, and bits, that must be transmitted. Accordingly, high-resolution images, may be downscaled and transmitted at CIF resolution, i.e., 352x288 pixels, when 56k modems are used to receive the transmitted image, or may be transmitted using CCIR quality, i.e., 426x576 pixels over a higher speed data link. Transmission of images having higher resolutions is achievable using for example, Digital Subscriber Lines (DSL) or cable modems.

[0038] Figure 6 illustrates another embodiment of spatial scalability that includes motion compensation performed on the high-resolution images. More specifically, spatial layer 520, in a manner similar to that previously discussed with regard to Figure 5, and spatial layer frames 522, 523, 524, etc., are motion compensated by using video information items contained in the high-resolution enhancement layer frames. In this illustrative example, motion compensation is performed on frame 526, for example, by using video data from previous frames, 522, 524, and subsequent frames, 528 (not shown). Motion compensation of encoded enhancement layer data is more fully described in the related patent applications and need not be discussed further herein.

[0039] Figure 7 illustrates another aspect of the present invention, wherein image 2' is encoded such that base layer 210' includes lower frequency elements of a video image and SNR enhancement layer 240' and spatial enhancement layer 530 include higher frequency elements of a video image. In this embodiment, low frequency elements are given a higher transmission priority than to those of higher frequencies, as base layer 210' is always transmitted. Thus, when a bit rate is below a known threshold, e.g., R_T, the sequence is displayed at a lower resolution. And, at higher bit rates, the sequence is displayed at the higher resolution.

[0040] Figure 8 illustrates a hybrid-spatial-temporal-SNR FGS encoding in accordance with the principles of the encoder illustrated in Figure 4b. In this aspect of the invention, original video signal 2' is downscaled into downscaled base layer 210', containing frames 212, 214. A downscaled temporal enhancement layer 230' is generated from the data contained in frames 212, 214. A downscaled SNR enhancement layer 242' is generated from the original downscaled image and base layer 210' frame date. Spatial layer 540 is the

generated from an video image reconstructed by upscaling video data information contained in corresponding frames, i.e., 212'/242', 214'/244', of base layer 210' and 240' and original video image 2'. A temporal enhancement layer 810 is then created from spatial enhancement layer frames, illustratively represented as 542, 544. Generation of temporal layer frames 812, 813, 814, etc., is similar to the generation of temporal layer frames 232', etc., and need not be discussed in detail herein.

[0041] Figure 9a illustrates an exemplary transmission system 900a utilizing the principles of the present invention. Video data is provided by video frame source 2 to video encoding unit 910. Video encoding unit 910 includes encoder 400a or 400b, similar to that illustrated in either Figures 4a or 4b, respectively. Video encoded data is then stored in encoder buffer 914. The encoded data is then provided to server 916 which transmits a encoded base layer and portions of the encoded enhancement layer data, individually or in combination, over data network 6. At receiving system 917, the received data frames are stored in decoder buffer 918 and provided to video decoder 920. Video decoder 920 extracts and decodes the received information. The decoded information items are next presented on video display 922 or may be stored on a, not shown, video recording device, such as an analog video recorder, a digital video recorder, writeable optical medium. Figure 9b illustrates a second exemplary transmission system utilizing the principles of the present invention. In this exemplary transmission system, the encoded data stored in encoder buffer 914 is provided to high bandwidth network 930. The high bandwidth transmitted data is then held in proxy system 932 for transmission over low bandwidth network 936 to decoder 917. At server 916 or proxy 932 a determination of the transmission bit rate may be determined based on the available bandwidth

[0042] Figure 10 shows an exemplary embodiment of a system 1000 which may be used for implementing the principles of the present invention. System 1000 may represent a television, a set-top box, a desktop, laptop or palmtop computer, a personal digital assistant (PDA), a video/image storage device such as a video cassette recorder (VCR), a digital video recorder (DVR), a TiVO device, etc., as well as portions or combinations of these and other devices. System 1000 includes one or more video/image sources 2, one or more input/output devices 1002, a processor 1003 and a memory 1004. The video/image source(s) 2 may represent, e.g., a television receiver, a VCR or other video/image storage device. The source(s) 2 may alternatively represent one or more network connections for receiving video from a server or servers over, e.g., a global computer communications network such as the Internet, a wide area network, a metropolitan area network, a local area network, a terrestrial broadcast system, a cable network, a satellite network, a wireless network, or a telephone network, as well as portions or combinations of these and other types of networks.

[0043] The input/output devices 1002, processor 1003 and memory 1004 may communicate over a communication medium 6. The communication medium 6 may represent, e.g., a bus, a communication network, one or more internal connections of a circuit, circuit card or other device, as well as portions and combinations of these and other communication media. Input video data from the source(s) 2 is processed in accordance with one or more software programs stored in memory 1004 and executed by processor 1003 in order to generate output video/images supplied to a display device 1006.

[0044] In a preferred embodiment, the coding and decoding employing the principles of the present invention may be implemented by computer readable code executed by the system. The code may be stored in the memory 1004 or read/downloaded from a memory

medium such as a CD-ROM or floppy disk. In other embodiments, hardware circuitry may be used in place of, or in combination with, software instructions to implement the invention. For example, the elements illustrated herein may also be implemented as discrete hardware elements.

[0045] Although the invention has been described in a preferred form with a certain degree of particularity, it is understood that the present disclosure of the preferred form has been made only by way of example, and that numerous changes in the details of construction and combination and arrangement of parts may be made without departing from the spirit and scope of the invention as hereinafter claimed. It is intended that the patent shall cover by suitable expression in the appended claims, whatever features of patentable novelty exist in the invention disclosed. Furthermore, it would be understood that reference to enhancement layer includes individual quality (SNR), temporal and spatial layers, in addition to combinations of SNR, temporal and spatial enhancement layers.